Calibración de modelos gravitacionales acotados en origen, para predecir variaciones en el total atraído de movilidad laboral

Report de recerca Nº 4

Jorge Cerda Troncoso

Abril 2010

Problema de investigación

El tradicional problema de gravitación planteado por Newton, y utilizado por Wilson para representar problemas de interacción espacial entre distintos territorios, ha tenido un gran desarrollo en distintas áreas. Los más importantes han sido en la economía (específicamente en la economía locacional), y también en el análisis y modelación de transporte (modelos de distribución de viajes). Es en estos últimos que surge el problema de la presente investigación.

La distribución de viajes se representa por una matriz, donde las filas representan las zonas desde donde salen los viajes, y las columnas las zonas a donde llegan los viajes. La celda de la matriz representa la cantidad de viajes que salen de una zona específica y que llegan a otra zona específica. A continuación se presenta un esquema explicativo.

$$
\begin{array}{ccccccccccc}
Zona.Destino & 1 & 2 & . & . & j & . & . & n & & \sum_j V_{ij} \\
Zona.Origen & & & & & & & & & & \\
1 & V_{11} & V_{12} & . & . & V_{1j} & . & . & V_{1n} & & O_1 \\
2 & V_{21} & V_{22} & . & . & V_{2j} & . & . & V_{2n} & & O_2 \\
. & . & . & & & . & & & . & & \\
. & . & . & & & . & & & . & & \\
i & V_{i1} & V_{i2} & . & . & V_{ij} & . & . & V_{in} & & O_i \\
. & . & . & & & . & & & . & & \\
. & . & . & & & . & & & . & & \\
n & V_{n1} & V_{n2} & . & . & V_{nj} & . & . & V_{nn} & & O_n \\
\sum_i V_{ij} & D_1 & D_2 & . & . & D_j & . & . & D_n & & \sum_{ij} V_{ij} = V
\end{array}
$$

El tradicional y más general modelo de distribución de viajes gravitacional, ya sea de su derivación física como entrópica, es el que se conoce como modelo gravitacional doblemente acotado, en el que se resuelve el problema de distribuir viajes (al interior de la matriz), siendo conocidos tanto el total de viajes que sale de cada zona, como el total de viajes que llega a cada zona. Lo único que se desconoce es el cómo los viajes se distribuyen, cumpliendo con las restricciones de origen (cuantos viajes salen) y de destino (cuantos viajes llegan). Cabe mencionar que las posibles matrices que cumplen las restricciones de orígenes y destinos suelen ser muchas (no hay solución única), por lo que se podría resolver como un problema tradicional de optimización de costos (mínimo costo total del sistema), pero el problema del transporte, por ser un problema de decisiones humanas, se reconoce más bien probabilístico que determinístico. De lo anterior es que surge el enfoque entrópico como método de solución. Este enfoque plantea que de todos los posibles estados del sistema (soluciones de la matriz), el más probable es el de mayor desorden del sistema (máxima entropía), el que se obtiene maximizando una función de entropía conocida. Este enfoque, sus ecuaciones finales, y sus métodos de resolución, son los más utilizados en la temática de la modelación agregada de demanda de transporte.

A continuación se presenta la ecuación de este modelo

$$V_{ij} = A_i O_i B_j D_j e^{-(\beta * d_{ij})}$$

donde

Vij	: viajes que van de la zona i a la zona j
Oi	: total de viajes que salen de la zona i
Dj	: total de viajes que llegan a la zona j
dij	: medida de separación entre la zona i y la zona j
Ai, Bj	: factores de balanceo, que se calculan internamente
β	: coeficiente de la función de fricción espacial

De la ecuación se puede observar que existen dos factores de balanceo, con los cuales siempre se logra cumplir las restricciones de totales originados y atraídos por zonas. Pero existe un solo coeficiente que se debe calibrar con una matriz observada, que corresponde al coeficiente que participa en la función de fricción del espacio (β). La función de fricción antes mencionada ha tenido una evolución en el tiempo, pasando de funciones de potencia inversa de la distancia, a funciones exponencial de costo generalizado de viaje.

Existen otras versiones más simples de este modelo, que corresponde a los modelos simplemente acotados, en los que solo se saben o los viajes que sale o los viajes que llegan a los distintos territorios. Considerando el modelo acotado a origen, es decir, se saben sólo los viajes que salen, el problema se resuelve con una estructura matemática similar a la anterior, considerando una masa de atracción en el destino. Si bien en este caso no se tiene una restricción de cantidad en el destino, las distribuciones están fuertemente condicionados por las proporcionen que tengan los atractivos, por lo que indirectamente se está resolviendo el problema de cuantos viajes llegan según la variable de atractivo que se considere.

La ecuación de estos modelos es la siguiente

$$V_{ij} = O_i B_j e^{-(\beta * d_{ij})}$$

donde

Vij	: viajes que van de la zona i a la zona j
Oi	: total de viajes que salen de la zona i
dij	: medida de separación entre la zona i y la zona j
Bj	: atractivo de la zona j
β	: coeficiente de la función de fricción espacial

En ciertos casos de análisis territorial, lo que se busca no es predecir el número absoluto de viajes que llegarán a una zona, sino más bien la proporción de los viajes que llegaran a una determinada zona, a partir de una zona de origen. Este “comportamiento” de la proporción será independiente del total de viajes que salgan de la zona, y se puede interpretar como una probabilidad de destino (espacial) que rige a cualquier viajero que salga de dicha zona.

El problema específico de investigación es el poder determinar una estructura de modelación acotada a origen, que prediga proporciones de llegada, y que sea sensible al cambio en los atractivos de los destinos, al cambio en la función de fricción, y que además se pueda calibrar con una matriz observada.

Objetivos y metodología

El objetivo del presente trabajo es formular y calibrar un modelo de distribución simplemente acotado en origen, de proporciones de viajes, que cumpla con las condiciones de ser sensible a los cambios en los pesos atractivos, y en las medidas de costos entre zonas.

Analizando la estructura del modelo acotado a origen, se puede construir un modelo que entregue como resultado la proporción de viajes que salen de una zona, y que se llegan a todos las zonas. La estructura matemática sería la siguiente:

$$P_{ij} = \frac{V_{ij}}{O_i} = B_j e^{-(\beta * d_{ij})}$$

donde

Pij	: proporción de viajes que sale de i y llegan a j
Vij	: viajes que van de la zona i a la zona j
Oi	: total de viajes que salen de la zona i
dij	: medida de separación entre la zona i y la zona j
Bj	: atractivo de la zona j
β	: coeficiente de la función de fricción espacial

El modelo anterior, si bien cumple con los requerido en el sentido de predecir proporciones, y permitir alteraciones en el variable de atractivo o en la variable de distancia entre las zonas, no permite calibrar ningún parámetro para el atractivo (cosa que si hace para la fricción). Además no asegura que la suma de todas las proporciones que salen de una zona sea 1. Por lo anterior, se definió la siguiente estructura de gravitación (de tipo potencial estandarizado).

$$P_{ij} = \frac{B_j^{\alpha j} * e^{-(\beta * d_{ij})}}{\sum_k B_k^{\alpha k} * e^{-(\beta * dik)}}$$

donde

Pij	: proporción de viajes que sale de i y llegan a j
Bj	: atractivo de la zona j
dij	: medida de separación entre la zona i y la zona j
αj	: coeficiente del atractivo de la zona j
β	: coeficiente de la función de fricción espacial

Esta formulación asegura que la suma de las proporciones, para una zona determinada, sume 1, además de permitir la calibración del coeficiente de la función de fricción espacial (único para todas las zonas), y coeficientes diferenciados del atractivo de la zona de destino (lo que permite reflejar un peso diferenciado propio de cada destino).

Este modelo se calibró a nivel de protosistemas (agregación funcional de municipios) de Cataluña (195), con la matriz de movilidad obligada del censo 2001 (POR-LTL), utilizando el método de minimizar la suma de los errores al cuadrado. También se consideraron dos estructuras matemáticas de la función de fricción espacial. Los valores de atractivo de las zonas fueron LTL totales, y los costos de interacción fue una matriz de tiempos de viajes obtenidos de la Autoridad Metropolitana del Transporte (ATM) a partir del modelo SIMCAT, para el año 2001.

El método para medir el ajuste de los modelos de distribución, en general presentan muchos problemas, ya que las diferencias que se produce en una celda específica no son detectadas, pues los métodos (por ejemplo el de Hymann ampliamente usado en transporte) generan indicadores agregados de error, con lo que se compensan los valores de las celdas.

En este estudio se aplicaron tres métodos de medición del ajuste, los que se presentan a continuación:

1.- Error cuadrático total.
2.- Coeficiente de correlación lineal entre los valores celda a celda de la matriz observada y la matriz modelada.
3.- Coeficiente de determinación, coeficiente de regresión, y error estándar de una regresión uni-variada, sin intercepto, donde los valores (celda a celda) de la matriz observada es la variable dependiente (y), y los valores de la matriz modelada son las variables independiente (x).

La ventaja o pertinencia de cada uno de estos indicadores se presentarán en los resultados.

Resultados

Los resultados de la calibración del modelo corresponden a los valores de los coeficientes, de las distintas estructuras modeladas, y los valores de ajuste general del modelo.

Las estructuras matemáticas alternativas que se evaluaron fueron las siguientes:

Potencia inversa

$$P_{ij} = \frac{LTL_j^{\alpha j} * Tij^{-\beta}}{\sum_k LTL_k^{\alpha k} * Tik^{-\beta}}$$

Exponencial inversa

$$P_{ij} = \frac{LTL_j^{\alpha j} * e^{-(\beta * T_{ij})}}{\sum_k LTL_k^{\alpha k} * e^{-(\beta * Tik)}}$$

A continuación se presentan los coeficientes obtenidos.

Tabla 1.- Resultados de la calibración de los modelos

Coeficientes	Función de fricción de potencia inversa	Función de fricción exponencial inversa
β	**-5,81**	**-13,67**
αj		
Valor medio	**0,90**	**0,81**
valor mínimo	-2,78	-8,40
valor máximo	2,26	3,00
Desviación estandar	0,87	1,48
CV	0,97	1,82

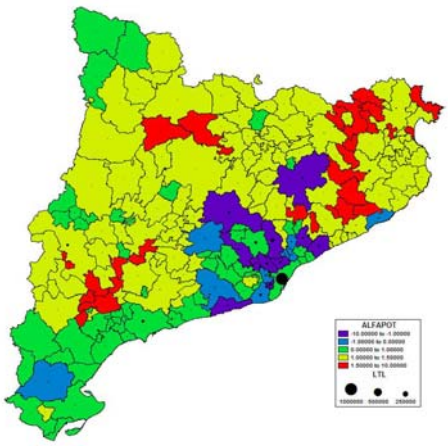
De la tabla 1 se observa que:

1.- El signo del coeficiente beta es coherente con la teoría para los dos modelos calibrados. La diferencia en la magnitud de los valores de estos coeficientes se debe a que la función exponencial presenta una mayor inercia de crecimiento en relación a la potencia, por lo que requiere de un mayor coeficiente para igual los valores con la función de potencia inversa.

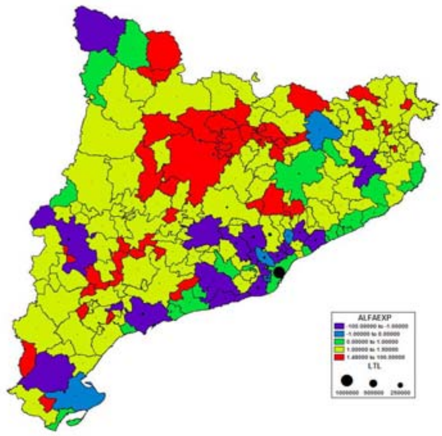
2.- El coeficiente alfa, que corresponde a la potencia del atractivo puede tener valore positivos y negativos. El valor de referencia para este coeficiente es el valor alfa=1, pues valores menores de 1 significa que la magnitud del atractivo se disminuye para lograr el ajuste del modelo (los valores negativos son disminuciones aún más significativas siendo asintóticas a cero). Cuando el valor de alfa es mayor que 1, el atractivo real se aumenta por efecto de la potencia. De los resultados se observa que en promedio los valores de alfa son menores de 1. El modelo de potencia inversa obtiene coeficientes alfa con menor dispersión estadística que el modelo exponencial. A continuación se muestra la distribución espacial de los valores de alfa obtenidos.

Figura 1.- Distribución espacial del coeficiente alfa calibrado por modelo

Modelo de potencia inversa



Modelo exponencial inversa



Los colores amarillo y rojo muestran valores de alfa mayores de 1, mientras que los colores verde, azul, y lila, muestran valores menores de 1.

En general los valores mayores de 1 esta al interior de Cataluña, mientras que los valores menores de 1 se encuentran en el borde costero y grandes municipios interiores (Girona, Lleida). Existen diferencias puntuales en los modelos, pero en general siguen el mismo patrón antes mencionado.

A continuación se muestran los ajustes logrados por los distintos modelos.

Tabla 2.- Ajuste de calibración de los modelos

Ajuste	Función de fricción de potencia inversa	Función de fricción exponencial inversa
Error cuadrático total	24,52	33,10
Coef. Determinación (R2)	**0,73**	**0,63**
Correlación lineal	0,86	0,80
Beta regresión univariada	**0,91**	**0,84**
Error estándar regresión	0,02	0,03

La tabla muestra distintos indicadores de ajuste, los que se explican y analizan a continuación:

1.- El error cuadrático total es la cantidad que minimiza el método de mínimos cuadrados. Los valores que se observan son los mínimos obtenidos para los distintos modelos. Estos valores son comparables, por lo que se puede decir que el modelo de potencia inversa es más eficiente en el ajuste global que el modelo exponencial, pues logra un menor mínimo.

2.- El coeficiente de determinación surge de calibrar una regresión lineal uni-variada, donde la variable dependiente (y) es la proporción observada por celda, y la variable independiente (x) es la proporción modelada para la misma celda. No se considera constante de regresión. Aclarado esto, vemos que el modelo con potencia inversa se acerca en mayor porcentaje (10% más) a las proporciones observadas que el modelo exponencial.

3.- El coeficiente de correlación lineal entre los valores observados y modelados también indica que el modelo de potencia inversa es mejor que el exponencial, aunque este indicador no refleja la igualdad de los valores, sino más bien la proporcionalidad de los mismos.

4.- El coeficiente beta de la regresión uni-variada sin constante, muestra la relación (promedio) entre los valores modelados y los observados. Si este coeficiente obtuviese un valor igual a 1, el ajuste sería perfecto, pues significa que el valor modelado es igual al valor observado. En nuestro caso, es el modelo de potencia inversa el que mejor se ajusta (valor 0,91 contra un 0,84 del otro modelo).

5.- El error estándar de la predicción del modelo de regresión uni-variada sin constante indica en cuanto se equivoca (de forma estándar) el modelo. El modelo de potencia inversa se equivoca en un 2% en la proporción, mientras que el modelo exponencial se equivoca en un 3% de proporción. Estos valores son coherentes con los valores del coeficiente de determinación (analizados en el punto 2), que indican el porcentaje en que se acierta.

Conclusiones

Finalmente se pudo estructurar un modelo que cumpliese con los requerimientos del fenómeno en estudio, es decir, que modelara proporciones (comportamiento) y no valores absolutos, que se pudiese calibrar con una matriz observada, y que pudiese sensibilizar el efecto en variaciones en el atractivo y/o en la matriz de costos de interacción espacial.

El calibrar coeficientes distintos según el territorio que atrae, primero logra un mejor ajuste del modelo, y segundo refleja las diferencias territoriales propias de la diversidad de cada uno de las zonas analizadas.

Los ajustes son relativamente buenos, lo que significa que se replica bien la realidad observada el año 2001.

El mejor modelo, y el que en definitiva se utilizará en la solución del problema en cuestión, surge de la coincidencia de todos los criterios analizados, y corresponde al modelo de función de fricción de potencia inversa.